5.1 & 5.2 Frequency Tables, Histograms & Freq Polygons - Key

Frequency Tables, Histograms, & Frequency Polygons [5.1 & 5.2]

Are you ready for some STATS???? Let's start by exploring the similarities & differences between 2 sets of data.

This is a bit too much info for us. Let's focus on the 1st two lines

Measured Lifespans of 30 Car Batteries (years)									
Brand X					Brand Y				
5.1	7.3	6.9	4.7	5.0	5.4	6.3	4.8	5.9	5.5
6.2	6.4	5.5	5.7	6.8	4.7	6.0	4.5	6.6	6.0
6.0	4.8	4.1	5.2	8.1	5.0	6.5	5.8	5.4	5.1
6.3	7.5	5.0	5.7	8.2	5.7	6.8	5.6	4.9	6.1
3.3	3.1	4.3	5.9	6.6	4.9	5.7	6.2	7.0	5.8
5.8	6.4	6.1	4.6	5.7	6.8	5.9	5.3	5.6	5.9

Mean	Median	Mode
"the average" add all the values and divide by how many values there are.	"the middle" line them up and cross off from both sides	"Most Common" → Occurs most often
x: (5.1 + 7.3 + 6.9 + 4.7 + 5.0 + 6.2 + 6.4 + 5.5 + 5.7 + 6.8) = 59.6 59.6 ÷ 10 = 6.0 $\boxed{x: 6.0}$ y: 55.7 ÷ 10 = 5.57 $\boxed{y: 5.6}$	~~4.7~~, ~~5.0~~, ~~5.1~~, ~~5.5~~, 5.7, 6.2, ~~6.4~~, ~~6.8~~, ~~6.9~~, ~~7.3~~ If only 1 number → that's it! If not: $\frac{5.7+6.2}{2} = 5.95 = \boxed{6.0}$ $\boxed{y = 5.7}$	x: Look at chart. No values occurs more than once $\boxed{\therefore \text{ x: none}}$ y: 6.0 occurs 2x $\boxed{y = 6.0}$

Range: Highest # − Lowest #

x: 7.3 − 4.7 = 2.6
$\boxed{x = 2.6}$

y: 6.6 − 4.5 = 2.1
$\boxed{y = 2.1}$

Maximum Water Flow Rates for the Red River, from 1950 to 1999, Measured at Redwood Bridge*									
Year	Flow Rate (m^3/s)	Year	Flow Rate (m^3/s)	Year	Flow Rate (m^3/s)	Year	Flow Rate (m^3/s)	Year	Flow Rate (m^3/s)
1950	3058	1960	1965	1970	2280	1980	881	1990	396
1951	1065	1961	481	1971	1526	1981	159	1991	280
1952	1008	1962	1688	1972	1589	1982	1458	1992	1399
1953	357	1963	660	1973	530	1983	1393	1993	946
1954	524	1964	1002	1974	2718	1984	1048	1994	1121
1955	1521	1965	1809	1975	1671	1985	991	1995	1877
1956	1974	1966	2498	1976	1807	1986	1812	1996	3058
1957	654	1967	1727	1977	187	1987	2339	1997	4587
1958	524	1968	510	1978	1750	1988	564	1998	1557
1959	991	1969	2209	1979	3030	1989	1390	1999	2180

Example 1: Determine the water flow rate that is associated with serious flooding by creating a frequency distribution. (table or graph)

* a frequency distribution should have between 5 and 12 intervals

To determine my intervals, I look at the Range.

Highest: 4587 Range: 4587 - 159 = 4428

Lowest: 159

I want 10 intervals: 4428 ÷ 10 = 442.8 → 450 → 500

easier # to work with

Frequency Distribution Table

Intervals

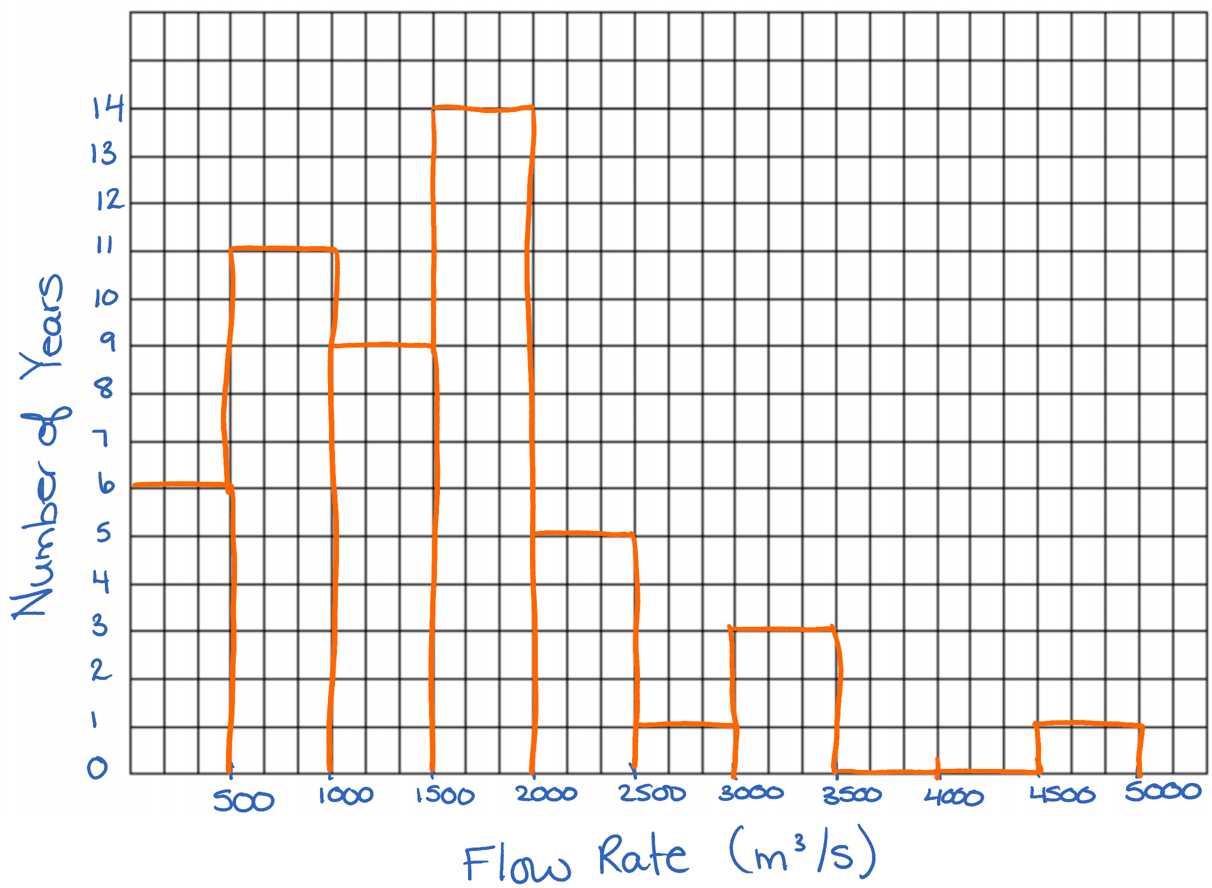
Flow Rate (m^3/s)	Tally	Frequency (# of years)	Midpoint (Polygon)
0-500	~~||||~~ |	6	250
500-1000	~~||||~~ ~~||||~~ |	11	750
1000-1500	~~||||~~ ||||	9	1250
1500-2000	~~||||~~ ~~||||~~ ||||	14	1750
2000-2500	~~||||~~	5	2250
2500-3000	|	1	2750
3000-3500	|||	3	3250
3500-4000		0	3750
4000-4500		0	4250
4500-5000	|	1	4750

Using the same information from above, create a Histogram.

Things to remember when creating Histograms

- graph of a frequency distribution.
- equal intervals marked on horizontal axis
- frequencies on vertical axis
- No spaces between bars
- Must have a title and labelled axis

River Flow Rates (1950-1999)

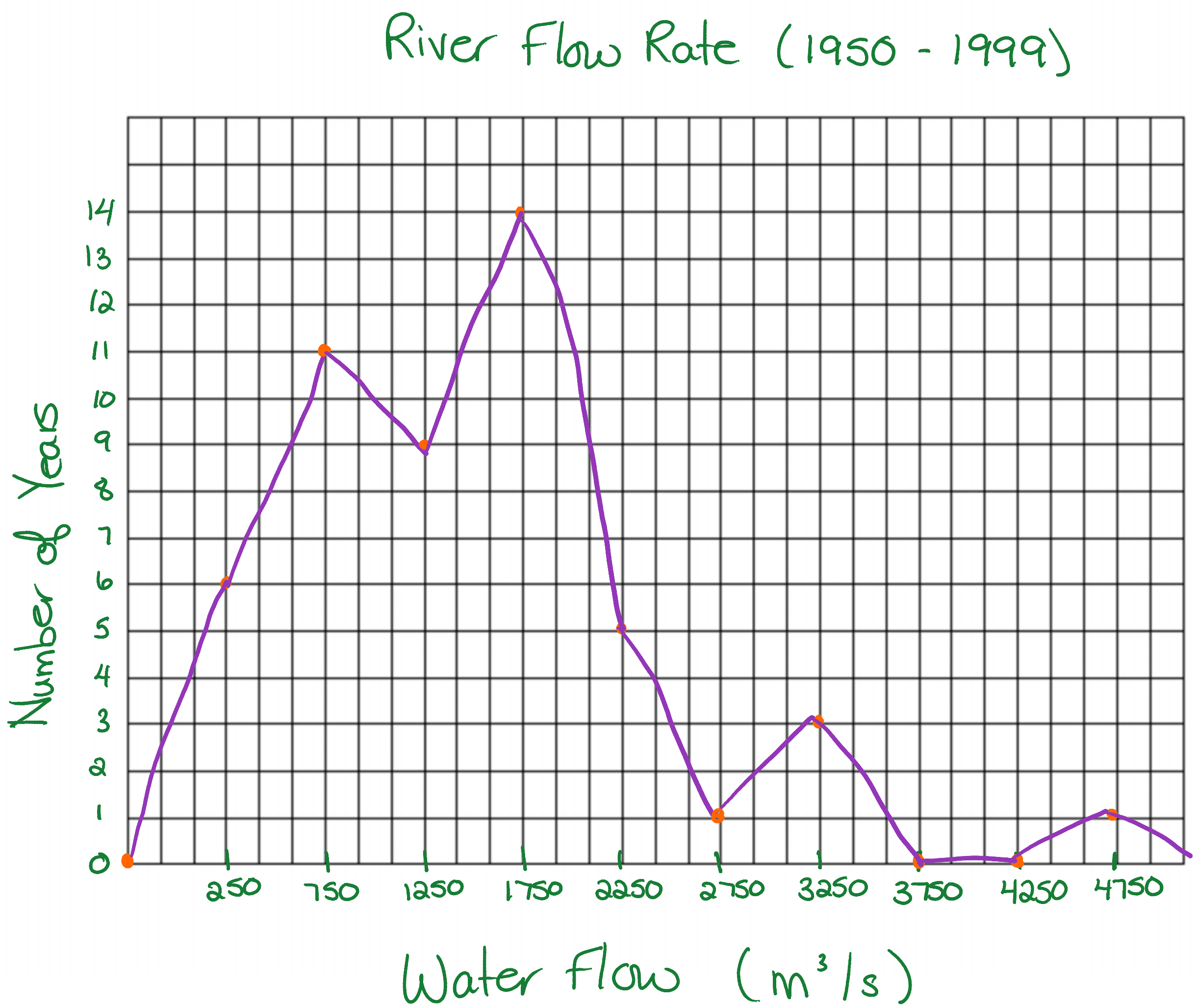


* Evenly distributed: 32 squares ÷ 10 intervals = 3.2

Look what else we can do with the same information… a <u>Frequency Polygon</u>!

A Quick How To: the Frequency Polygon

- Use the midpoints of the intervals
- Connect points using straight lines

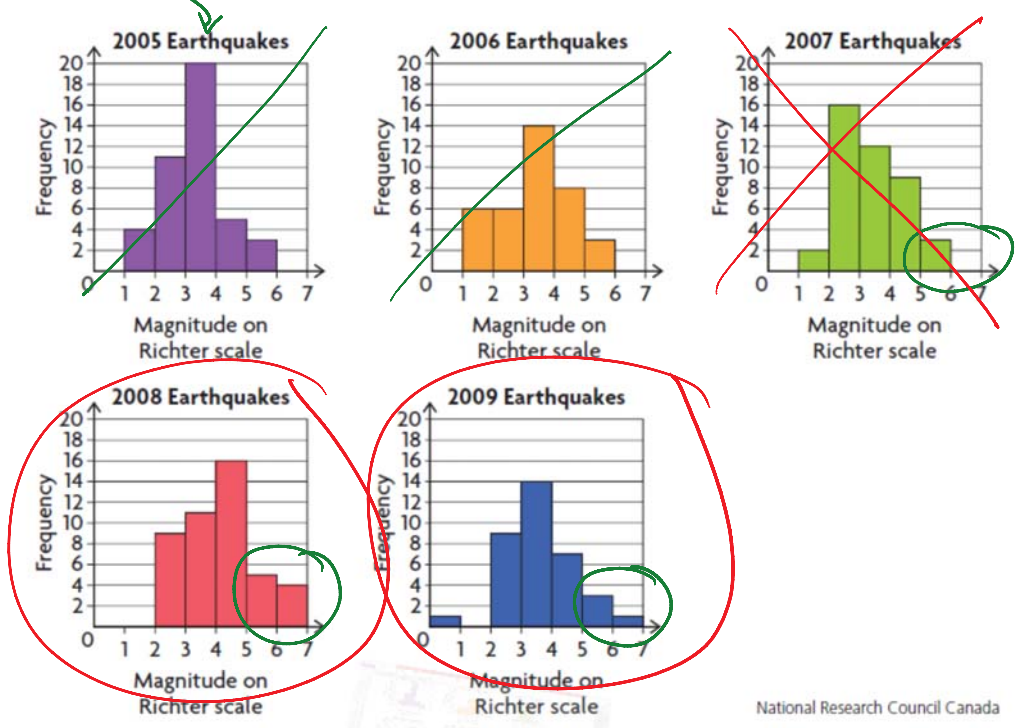


Example 2:

Which of these years could have had the most damage from earthquakes?

Understanding the Richter Scale*	
Magnitude	**Effects**
less than 3.0	recorded by seismographs; not felt
3.0–3.9	feels like a passing truck; no damage
4.0–4.9	felt by nearly everyone; movement of unstable objects
5.0–5.9	felt by all; considerable damage to weak buildings
6.0–6.9	difficult to stand; partial collapse of ordinary buildings
7.0–7.9	loss of life; destruction of ordinary buildings
more than 7.9	widespread loss of life and destruction

*Every unit increase on the Richter scale represents an earthquake 10 times more powerful. For example, an earthquake measuring 5.6 is 10 times more powerful than an earthquake measuring 4.6.

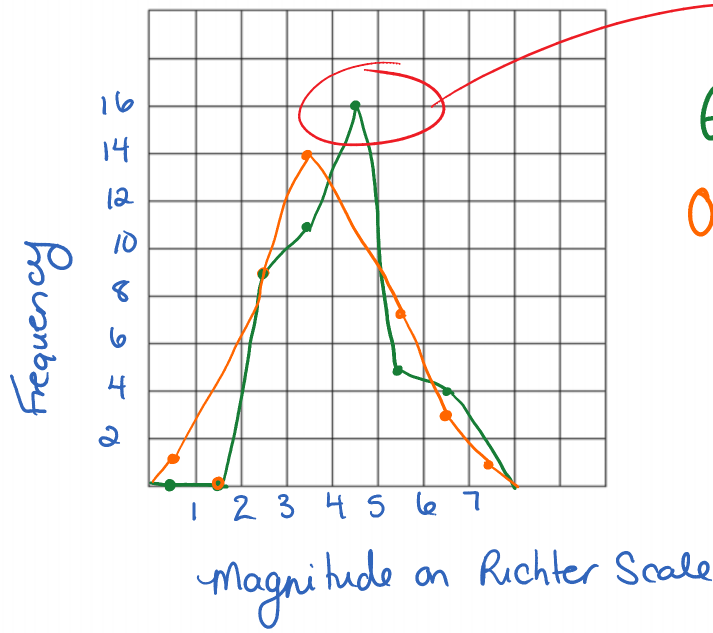


National Research Council Canada

Frequency Polygon:

Midpoints

Earthquakes 2008 + 2009



Most damage

Green 2008

Orange 2009

Homework: p. 211 # 1, 2

p. 222 # 4, 6, 8, 9

5.3 Standard Deviation - Key

Standard Deviation [5.3]

Standard deviation: a measure of the dispersion or scatter of data values in relation to the mean.

$$\sigma = \sqrt{\frac{\sum(x-\overline{x})^2}{n}}$$

σ (read as sigma – lower case): represents the standard deviation of the data (population)

Σ (read as sigma – upper case): summation operator

x: each data value

$\overline{x}$ (read as x bar): represents the mean of the data

n: the number of data values

$\frac{\text{adding all the numbers together}}{\text{divide by the \# of data values}}$

Example 1:

Brendan works part-time in the canteen at his local community centre. One of his tasks is to unload delivery trucks. He wondered about the accuracy of the mass measurements given on two cartons that contained sunflower seeds. He decided to measure the masses of the 20 bags in the two cartons. One carton contained 227 g bags, and the other carton contained 454 g bags.

x

Masses of 227 g Bags (g)			
228	220	233	227
230	227	221	229
224	235	224	231
226	232	218	218
229	232	236	223

Masses of 454 g Bags (g)			
458	445	457	458
452	457	445	452
463	455	451	460
455	453	456	459
451	455	456	450

$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$

How can measures of dispersion be used to determine if the accuracy of measurement is the same for both bag sizes?

① Start by finding the mean of your population

$\bar{x} = \frac{4543}{20} = 227.15g$

$n = 20$

$(228 - 227.15)^2$
$(0.85)^2$
$= 0.7225$

The standard deviation for the other bags was 4.4988...

* This means the 454g bag is more consistent because it has a smaller standard deviation.

②

x	$(x-\bar{x})^2$
228	0.7225
230	8.1225
224	9.9225
226	1.3225
229	3.4225
220	51.1225
227	0.0225
235	61.6225
232	23.5225
232	23.5225
233	34.2225
221	37.8225
224	9.9225
218	83.7225

x	$(x-\bar{x})^2$
236	78.3225
227	0.0225
229	3.4225
231	14.8225
218	83.7225
223	17.2225

③ Complete the formula by adding them all up

$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$

$\sigma = \sqrt{\frac{546.55}{20}}$

$\sigma = 5.22757...$

Example 2:

Angèle conducted a survey to determine the number of hours per week that Grade 11 males in her school play video games. She determined that the mean was 12.84 h, with a standard deviation of 2.16 h.

$\bar{x} = 12.84$

$\sigma = 2.16$

Janessa conducted a similar survey of Grade 11 females in her school. She organized her results in this frequency table. Compare the results of the two surveys.

Gaming Hours per Week for Grade 11 Females		Midpoint
Hours	Frequency	
3–5	7	4
5–7	11	6
7–9	16	8
9–11	19	10
11–13	12	12
13–15	5	14

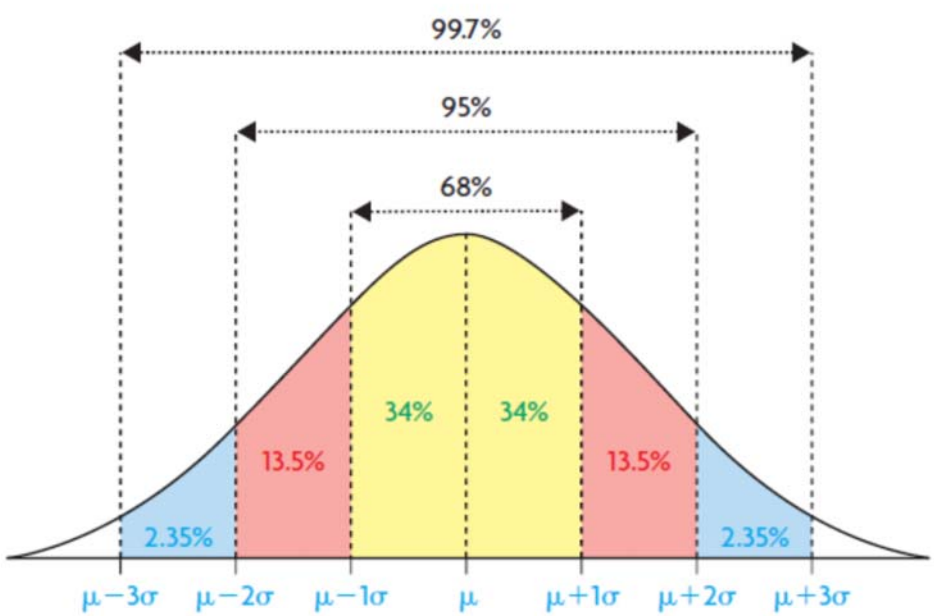
STAT
EDIT
L_1 → midpoint
L_2 → frequencies
STAT
CALC
1 - Var Stats (2^{nd} L_1, 2^{nd} L_2)
ENTER

5.4 The Normal Distribution - Key

The Normal Distribution [5.4]

Need to Know

- The properties of a normal distribution can be summarized as follows:
  - The graph is symmetrical. The mean, median, and mode are equal (or close) and fall at the line of symmetry.
  - The normal curve is shaped like a bell, peaking in the middle, sloping down toward the sides, and approaching zero at the extremes.
  - About 68% of the data is within one standard deviation of the mean.
  - About 95% of the data is within two standard deviations of the mean.
  - About 99.7% of the data is within three standard deviations of the mean.
  - The area under the curve can be considered as 1 unit, since it represents 100% of the data.



- Generally, measurements of living things (such as mass, height, and length) have a normal distribution.

Example 1:

Heidi is opening a new snowboard shop near a local ski resort. She knows that the recommended length of a snowboard is related to a person's height. Her research shows that most of the snowboarders who visit this resort are males, 20 to 39 years old. To ensure that she stocks the most popular snowboard lengths, she collects height data for 1000 Canadian men, 20 to 39 years old. How can she use the data to help her stock her store with boards that are the appropriate lengths?

n = 1000

When dealing with frequency we need midpoint

Midpoint	Height (in.)	Frequency
60.5	61 or shorter	3
61.5	61–62	4
62.5	62–63	10
	63–64	18
	64–65	30
	65–66	52
	66–67	64
	67–68	116
	68–69	128
69.5	69–70	147
	70–71	129
	71–72	115
	72–73	63
	73–74	53
	74–75	29
	75–76	20
	76–77	12
	77–78	5
78.5	taller than 78	2

↓ Data

425

① Find your mean → Using Graphing Calc + Cheat Sheet

$\bar{x} = 69.521$

$\sigma = 2.987$ (Standard Deviation)

② Find your median

median = 69.5 → Crossing out top + bottom until you reach the middle (of midpoints)

person 500/501

→ adding freq. to here gives us 425. Half of sample is 500. Person "500" has to be within this interval.

③ Find your mode

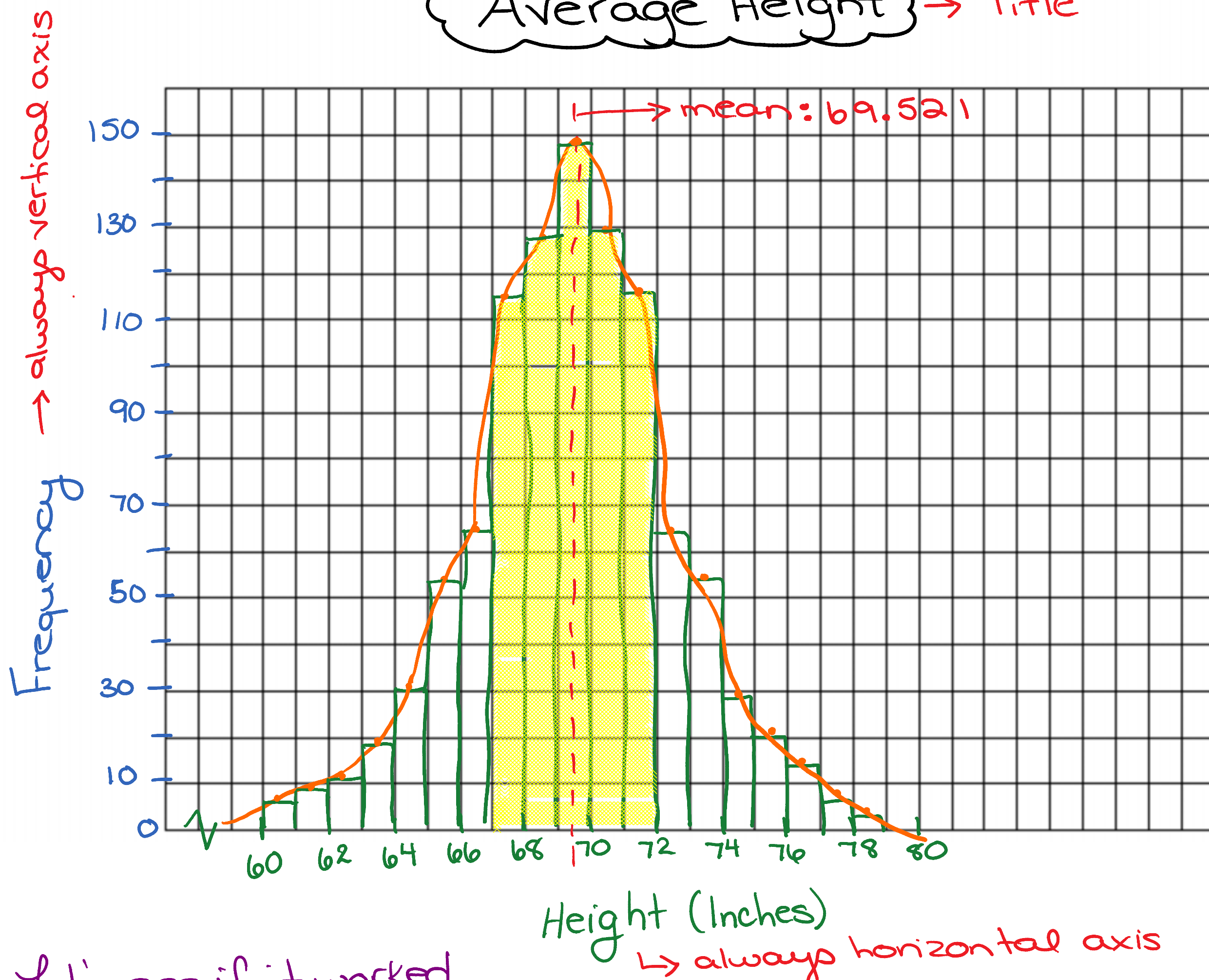
mode = 69.5 → looking at frequency, which interval has the most frequencies (highest #)

④ We can use this info to create a histogram (next page)

↳ Check notes from 5.1 + 5.2 for things to remember.

↳ Range: lowest freq = 2
highest freq = 147

Average Height → Title



Notice
- Bell curve
- Symm-etrical

Let's see if it worked

$\sigma = 2.987$

→ one SD from the mean

$69.521 - 2.987 = 66.5''$

$69.521 + 2.987 = 72.5''$

How many men's heights were between 67" – 73"

↳ $116 + 128 + 147 + 129 + 115 + 63$

$= 698$ (of our population)

If we looked at 1000 men

$\frac{698}{1000} = 69.8\% \rightarrow 70\%$ ✓

→ two SD from the mean

$69.521 - 2(2.987) = 63.5''$

$69.521 + 2(2.987) = 75.5''$

Between 64" – 76"

↳ $30 + 52 + 64 + 116 + 128 + 147 +$
$129 + 115 + 63 + 53 + 29 + 20$
$= 946$

$\therefore \frac{946}{1000} = 94.6\%$ ✓

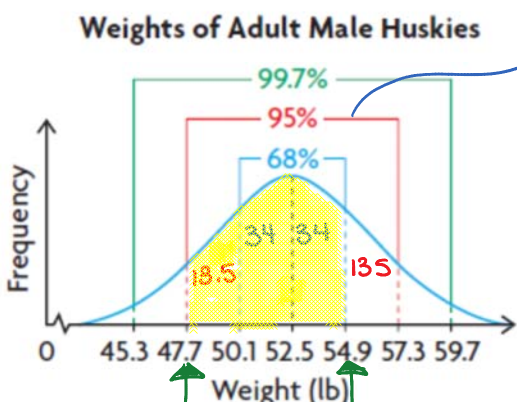
$\boxed{\text{∴ She wants 70\% of her boards to fit men between the heights of 66.5" and 72.5".}}$

Example 2:

Jim raises Siberian husky sled dogs at his kennel. He knows, from the data he has collected over the years, that the weights of adult male dogs are normally distributed, with a mean of 52.5 lb and a standard deviation of 2.4 lb. Jim used this information to sketch a normal curve, with

$\bar{x} = 52.5$ lbs
$\sigma = 2.4$ lbs

- 68% of the data within one standard deviation of the mean
- 95% of the data within two standard deviations of the mean
- 99.7% of the data within three standard deviations of the mean



$95\% - 65° = \frac{27\%}{2} \rightarrow 13.5\%$

What percent of adult male dogs at Jim's kennel would you expect to have a weight between 47.7 lb and 54.9 lb?

How many fall in that range?

$13.5 + 34 + 34 = 81.5\%$

$\boxed{81.5\%}$

Example 3:

Two baseball teams flew to the North American Indigenous Games. The members of each team had carry-on luggage for their sports equipment. The masses of the carry-on luggage were normally distributed, with the characteristics shown to the right.

Team	μ (kg)	σ (kg)
Men	6.35	1.04
Women	6.35	0.59

a) Sketch a graph to show the distribution of the masses of the luggage for each team.

b) The women's team won the championship. Each member received a medal and a souvenir baseball, with a combined mass of 1.18 kg, which they packed in their carry-on luggage. Sketch a graph that shows how the distribution of the masses of their carry-on luggage changed for the flight home.

5.5 Z-Scores - Old Key 1

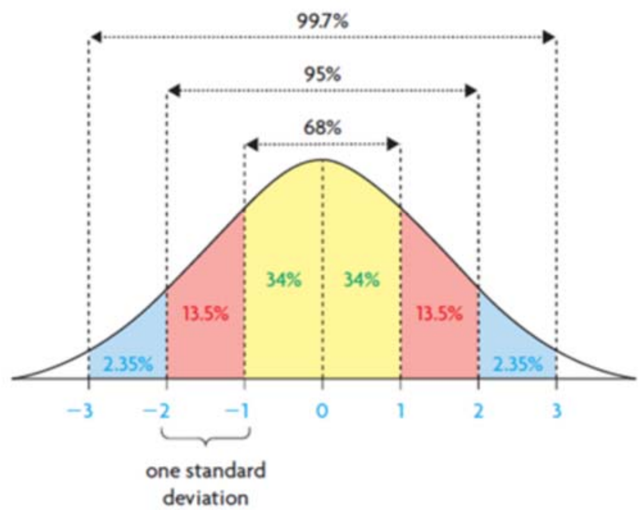
Z- Scores [5.5]

In Summary

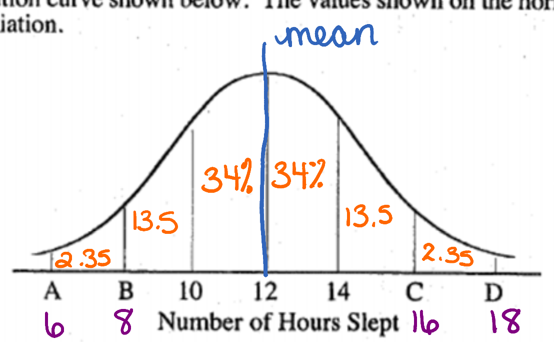
Key Ideas

- The standard normal distribution is a normal distribution with mean, μ, of 0 and a standard deviation, σ, of 1. The area under the curve of a normal distribution is 1.
- Z-scores can be used to compare data from different normally distributed sets by converting their distributions to the standard normal distribution.

μ = mu
σ = theta



A nurse records the number of hours an infant sleeps during a day. He then records the data on a normal distribution curve shown below. The values shown on the horizontal axis differ by one standard deviation.



a) What is the mean of the data?
12 hours

b) What is the standard deviation?
If each interval is 1 SD,
SD = 2 hours (14-12=2)

c) What are the values for A, B, C, and D?
6, 8, 16, 18

d) What percentage of a day, to the nearest hundredth, does the infant sleep:

i) between 12 and 14 h?
34%

ii) between 8 and 16 h?
2(13.5 + 34)
95%

~~iii) less than 6 h?~~

e) Why is it not possible at this time to determine the percentage of a day that the infant sleeps for less than 13 hours?

This model is limited to data values that are exactly 1, 2, or 3 standard deviations from the mean.

For ½ deviations... we need z-scores

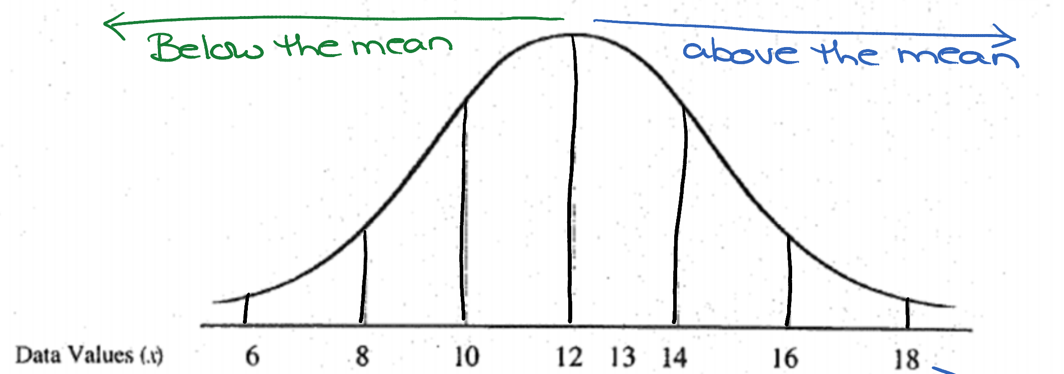
z–scores

A *z*-**score** for a data value describes the number of standard deviations above or below the mean.

Displaying z-scores on the Normal Curve

Consider Class Example #1 with a mean value of 12 and a standard deviation of 2.

- The data value 16 is 2 standard deviations above the mean and has a *z*-score of 2.
- The data value 6 is 3 standard deviations below the mean and has a *z*-score of -3.
- The data value 12 is 0 standard deviations away from the mean and has a *z*-score of 0.
- The data value 13 is ½ standard deviations above the mean and has a *z*-score of 0.5.



(mean - 3 stand.Dev) $\mu - 3\sigma$ $\mu - 2\sigma$ $\mu - \sigma$ μ $\mu + \sigma$ $\mu + 2\sigma$ $\mu + 3\sigma$

$\mu + 0.5\sigma$

z-scores: -3 -2 -1 0 0.5 1 2 3

z-score Formula

z-scores can be calculated using the formula

$$z = \frac{x - \mu}{\sigma}$$

where,
z is the z-score
x is the particular data value
μ is the mean
σ is the standard deviation

This formula is on the formula sheet

↳ Check: $z_{13} = \frac{13-12}{2} = 0.5$ ✓

Example 2: Determine at which location Hailey's run time was better when compared to club results.

Hailey and Serge belong to a running club in Vancouver. Part of their training involves a 200 m sprint. Below are normally distributed times for the 200 m sprint in Vancouver and on a recent trip to Lake Louise. At higher altitudes, run times improve.

We want to compare → use z scores

Distractor

Location	Altitude (m)	Club Mean Time: μ for 200 m (s)	Club Standard Deviation: σ (s)	Hailey's Run Time (s)	~~Serge's Run Time (s)~~
Vancouver	4	25.75	0.62	24.95	~~25.45~~
Lake Louise	1661	25.57	0.60	24.77	~~26.24~~

Beautiful! →

Vancouver

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{24.95 - 25.75}{0.62}$$

$$\boxed{z = -1.290\ldots}$$

Lake Louise

$$z = \frac{x - \mu}{\sigma}$$

$$z = \frac{24.77 - 25.57}{0.60}$$

$$\boxed{z = -1.333\ldots}$$

↳ She was better than the club on both runs, but she was best at Lake Louise.

Tony's midterm marks are shown below, together with the class mean and standard deviation for each subject. By calculating *z*-scores, determine in which subject Tony performed best relative to the rest of the class.

Subject	Tony's Mark	Mean Mark	Standard Deviation
Mathematics	74	68	12
Chemistry	79	73	14
Physics	68	66	11

$$Z_{74} = \frac{74-68}{12} = 0.50$$

$$Z_{79} = \frac{79-73}{14} = 0.43$$

$$Z_{68} = \frac{68-66}{11} = 0.18$$

Relative to the rest of the class, Tony performed best in math ☺ (highest score).

$x = 119$

$\mu = 100$

$\sigma = 15$

$z = \dfrac{x - \mu}{\sigma}$

$z = \dfrac{119 - 100}{15}$

$z = 1.2666 \rightarrow 1.27$

1.2 → z-score (vertical)

7 → "Top indicator"

Use z-score table on p. 592

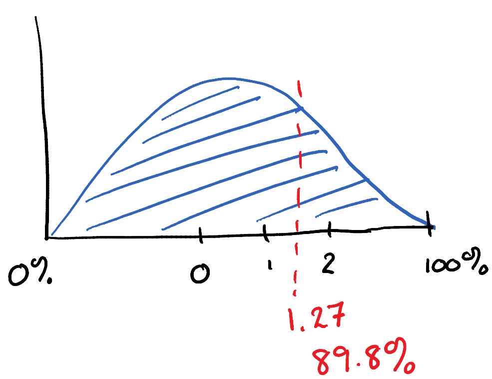
from chart: $\boxed{0.8980}$

89.8%

Example 4:

IQ tests are sometimes used to measure a person's intellectual capacity at a particular time. IQ scores are normally distributed, with a mean of 100 and a standard deviation of 15. If a person scores 119 on an IQ test, how does this score compare with the scores of the general population?

z-score!



An IQ score of 119 is greater than 89.8% of IQ scores in the population.

Example 5:

Athletes should replace their running shoes before the shoes lose their ability to absorb shock.

Running shoes lose their shock-absorption after a mean distance of 640 km, with a standard deviation of 160 km. Zack is an elite runner and wants to replace his shoes at a distance when only 25% of people would replace their shoes. At what distance should he replace his shoes?

Example 6:

The ABC Company produces bungee cords. When the manufacturing process is running well, the lengths of the bungee cords produced are normally distributed, with a mean of 45.2 cm and a standard deviation of 1.3 cm. Bungee cords that are shorter than 42.0 cm or longer than 48.0 cm are rejected by the quality control workers.

a) If 20 000 bungee cords are manufactured each day, how many bungee cords would you expect the quality control workers to reject?
b) What action might the company take as a result of these findings?

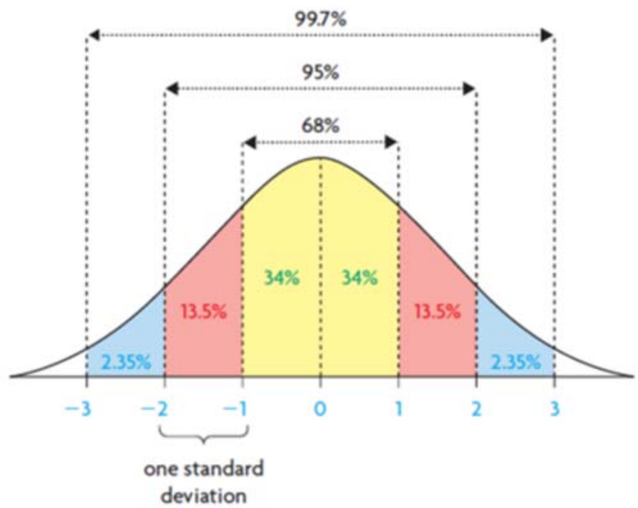
5.5 Z-Scores - Key

Z- Scores [5.5]

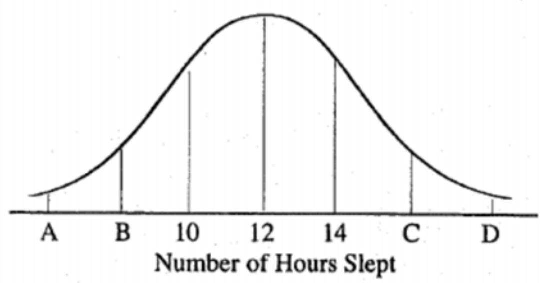
In Summary

Key Ideas

- The standard normal distribution is a normal distribution with mean, μ, of 0 and a standard deviation, σ, of 1. The area under the curve of a normal distribution is 1.
- *Z*-scores can be used to compare data from different normally distributed sets by converting their distributions to the standard normal distribution.



A nurse records the number of hours an infant sleeps during a day. He then records the data on a normal distribution curve shown below. The values shown on the horizontal axis differ by one standard deviation.



a) What is the mean of the data? **b)** What is the standard deviation?

c) What are the values for A, B, C, and D?

d) What percentage of a day, to the nearest hundredth, does the infant sleep:

i) between 12 and 14 h? **ii)** between 8 and 16 h? **iii)** less than 6 h?

e) Why is it not possible at this time to determine the percentage of a day that the infant sleeps for less than 13 hours?

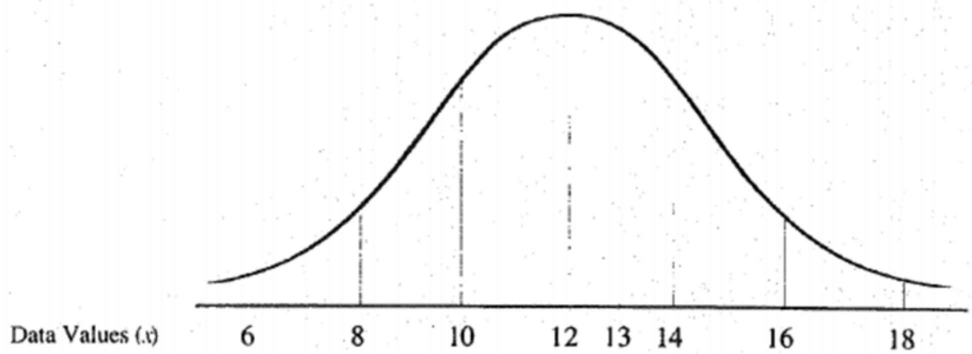
z–scores

A ***z*-score** for a data value describes the number of standard deviations above or below the mean.

Displaying z-scores on the Normal Curve

Consider Class Example #1 with a mean value of 12 and a standard deviation of 2.

- The data value 16 is ___ standard deviations above the mean and has a *z*-score of ___ .
- The data value 6 is ___ standard deviations below the mean and has a *z*-score of ___ .
- The data value 12 is ___ standard deviations away from the mean and has a *z*-score of ___ .
- The data value 13 is ____ standard deviations _______ the mean and has a *z*-score of ___ .



z-score *Formula*

z-scores can be calculated using the formula

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

This formula is on the formula sheet

where,
z is the z-score
x is the particular data value
μ is the mean
σ is the standard deviation

Example 2: Determine at which location Hailey's run time was better when compared to club results.

Hailey and Serge belong to a running club in Vancouver. Part of their training involves a 200 m sprint. Below are normally distributed times for the 200 m sprint in Vancouver and on a recent trip to Lake Louise. At higher altitudes, run times improve.

Location	Altitude (m)	Club Mean Time: μ for 200 m (s)	Club Standard Deviation: σ (s)	Hailey's Run Time (s)	Serge's Run Time (s)
Vancouver	4	25.75	0.62	24.95	25.45
Lake Louise	1661	25.57	0.60	24.77	26.24

Tony's midterm marks are shown below, together with the class mean and standard deviation for each subject. By calculating z-scores, determine in which subject Tony performed best relative to the rest of the class.

Subject	Tony's Mark	Mean Mark	Standard Deviation
Mathematics	74	68	12
Chemistry	79	73	14
Physics	68	66	11

The marks on a math exam at a university were found to have a mean of 52 with a standard deviation of 12. A professor who thought the exam was too difficult, decided to adjust the original marks by raising the mean to 65, while reducing the standard deviation to 10 and leaving the z-scores unchanged. What would the new mark be for a student who received an original mark of 34?

mark = 34

mean = 52

Stand. Dev. = 12

$$Z = \frac{x - \mu}{\sigma}$$

$$Z_{34} = \frac{34 - 52}{12}$$

$$z = -1.5$$

New Mean = 65

New St. Dev. = 10

$z = -1.5$

$$Z = \frac{x - \mu}{\sigma}$$

$$-1.5 = \frac{x - 65}{10}$$

$$-1.5 = x - 65$$

$\boxed{x = 50}$ ← New Mark

Example 5:

IQ tests are sometimes used to measure a person's intellectual capacity at a particular time. IQ scores are normally distributed, with a mean of 100 and a standard deviation of 15. If a person scores 119 on an IQ test, how does this score compare with the scores of the general population?

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319

Example 6:

Athletes should replace their running shoes before the shoes lose their ability to absorb shock.

Running shoes lose their shock-absorption after a mean distance of 640 km, with a standard deviation of 160 km. Zack is an elite runner and wants to replace his shoes at a distance when only 25% of people would replace their shoes. At what distance should he replace his shoes?

$\sigma = 160$

$\mu = 640$

$25\% \rightarrow 0.25\_\,\_$

0.2483 0.2514

$\boxed{-0.68 \qquad -0.67}$

-0.6 0.08 -0.6 0.07

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776

$\boxed{z = -0.675}$

$$Z = \frac{x - \mu}{\sigma}$$

$$\frac{-0.675}{1} = \frac{x - 640}{160}$$

→ Cross Multiply

$$x - 640 = 160(-0.675)$$

$$x = 160(-0.675) + 640$$

$$\boxed{x = 532 \text{ Km}}$$

Example 7:

The ABC Company produces bungee cords. When the manufacturing process is running well, the lengths of the bungee cords produced are normally distributed, with a mean of 45.2 cm and a standard deviation of 1.3 cm. Bungee cords that are shorter than 42.0 cm or longer than 48.0 cm are rejected by the quality control workers.

a) If 20 000 bungee cords are manufactured each day, how many bungee cords would you expect the quality control workers to reject?

b) What action might the company take as a result of these findings?

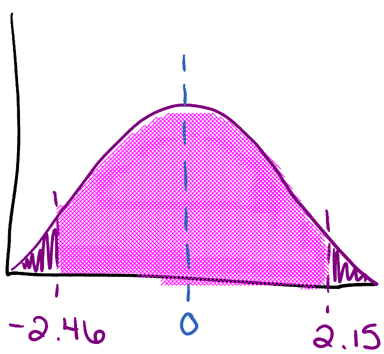
$\mu = 45.2$

$\sigma = 1.3$

$x = 42$ OR 48

$$Z = \frac{42 - 45.2}{1.3} \qquad Z = \frac{48 - 45.2}{1.3}$$

$$Z = -2.4615\ldots \qquad Z = 2.1538\ldots$$



0.0069   0.9842

$1 - 0.9842 = 0.0158$

(the area under the curve is 1, so the small portion is $1 - 0.9842$)

$0.0069 + 0.0158 = 0.0227$

$\downarrow$

2.27%

$0.0227 \times 20\,000 = \boxed{454 \text{ Bungee Cords}}$

5.6 Confidence Intervals - Key

Confidence Intervals [5.6]

In Summary

Key Ideas

- It is often impractical, if not impossible, to obtain data for a complete population. Instead, random samples of the population are taken, and the mean and standard deviation of the data are determined. This information is then used to make predictions about the population.
- When data approximates a normal distribution, a confidence interval indicates the range in which the mean of any sample of data of a given size would be expected to lie, with a stated level of confidence. This confidence interval can then be used to estimate the range of the mean for the population.
- Sample size, confidence level, and population size determine the size of the confidence interval for a given confidence level.

Need to Know

- A confidence interval is expressed as the survey or poll result, plus or minus the margin of error.
- The margin of error increases as the confidence level increases (with a constant sample size). The sample size that is needed also increases as the confidence level increases (with a constant margin of error).
- The sample size affects the margin of error. A larger sample results in a smaller margin of error, assuming that the same confidence level is required.

  For example:

  - A sample of 1000 is considered to be accurate to within $\pm 3.1\%$, 19 times out of 20.
  - A sample of 2000 is considered to be accurate to within $\pm 2.2\%$, 19 times out of 20.
  - A sample of 3000 is considered to be accurate to within $\pm 1.8\%$, 19 times out of 20.

The Effect of Sample Size on Margin of Error and Size of Confidence Interval

The 2011 Canadian federal election took place on May 2nd, 2011. A large number of opinion polls were conducted in the days leading up to the election.

(link: http://en.wikipedia.org/wiki/Opinion_polling_in_the_Canadian_federal_election,_2011)

The table below gives the results of four of these polls. The data represents the percent of the sample who would cast their vote for each party. Some of the results do not add up to 100% because votes for minor parties and independents are not included in the table. The results are accurate to the stated margin of error 19 times out of 20.

Polling Firm	Date of Poll	Number in Sample	Con	Lib	NDP	BQ	Green	Margin of Error
Harris Decima	Apr 27	1011	35%	22%	30%	5%	7%	±3.1%
Angus Reid	Apr 29	2197	37%	19%	33%	6%	4%	±2.2%
Nanos Research	Apr 30	1048	37.0%	22.7%	30.6%	5.5%	3.2%	±3.0%
EKOS Research	May 1	2690	33.9%	21.0%	31.2%	6.4%	6.0%	±1.8%

a) Look at the columns for **Number in Sample** and **Margin of Error**. How does the sample size affect the margin of error?

b) For each polling firm, calculate the 95% confidence interval for the percent of Canadian voters who would vote for the New Democratic Party (NDP) and complete the table. How does sample size affect the range of values in the confidence interval?

Polling Firm	95% Confidence Interval
Harris Decima	
Angus Reid	
Nanos Research	
EKOS Research	

c) For each polling firm, calculate the 95% confidence interval for the percent of Canadian voters who would vote for the Conservative Party and complete the table below.

Polling Firm	Confidence Interval Conservative	Confidence Interval NDP
Harris Decima		
Angus Reid		
Nanos Research		
EKOS Research		

i) In which of the polling firms is there no overlap of percents in the confidence intervals?

In this case, since there is no overlap, we can say with 95% confidence that the percent of Canadian voters who will vote Conservative is greater than the percent of Canadian voters who will vote NDP.

ii) In which of the polling firms is there an overlap of percents in the confidence intervals?

In these cases, since there is an overlap, we <u>cannot</u> say with 95% confidence that the percent of Canadian voters who will vote Conservative is greater than the percent of Canadian voters who will vote NDP.

d) In which of the polls can we say with 95% confidence that the NDP will receive a higher percent of the vote than the Liberal Party?

e) In which of the polls can we say with 95% confidence that the Bloc Quebecois will receive a higher percent of the vote than the Green Party?

Example 2:

In order to determine the mean mass of a type of chocolate bar produced at a factory, sampling is done and the following statement is made by the company.

" The mean mass of chocolate bars produced at our factory is 100.4 grams ± 0.4 grams. The results are accurate 18 times out of 20."

a) Determine the confidence level.

b) Determine the confidence interval.

c) State the margin of error.

d) Is it likely that the mean mass of chocolate bars produced at this factory is 99.5 grams? Explain.

e) Is it possible that the mean mass of chocolate bars produced at this factory is 99.5 grams? Explain.